Research Article

A Forecasting Approach Combining Self-Organizing Map with Support Vector Regression for Reservoir Inflow during Typhoon Periods

Gwo-Fong Lin,[1] Tsung-Chun Wang,[1] and Lu-Hsien Chen[2]

[1]*Department of Civil Engineering, National Taiwan University, Taipei 10617, Taiwan*
[2]*Department of Leisure Management, Taiwan Shoufu University, Tainan 72153, Taiwan*

Correspondence should be addressed to Tsung-Chun Wang; i246500@gmail.com

This study describes the development of a reservoir inflow forecasting model for typhoon events to improve short lead-time flood forecasting performance. To strengthen the forecasting ability of the original support vector machines (SVMs) model, the self-organizing map (SOM) is adopted to group inputs into different clusters in advance of the proposed SOM-SVM model. Two different input methods are proposed for the SVM-based forecasting method, namely, SOM-SVM1 and SOM-SVM2. The methods are applied to an actual reservoir watershed to determine the 1 to 3 h ahead inflow forecasts. For 1, 2, and 3 h ahead forecasts, improvements in mean coefficient of efficiency (MCE) due to the clusters obtained from SOM-SVM1 are 21.5%, 18.5%, and 23.0%, respectively. Furthermore, improvement in MCE for SOM-SVM2 is 20.9%, 21.2%, and 35.4%, respectively. Another SOM-SVM2 model increases the SOM-SVM1 model for 1, 2, and 3 h ahead forecasts obtained improvement increases of 0.33%, 2.25%, and 10.08%, respectively. These results show that the performance of the proposed model can provide improved forecasts of hourly inflow, especially in the proposed SOM-SVM2 model. In conclusion, the proposed model, which considers limit and higher related inputs instead of all inputs, can generate better forecasts in different clusters than are generated from the SOM process. The SOM-SVM2 model is recommended as an alternative to the original SVR (Support Vector Regression) model because of its accuracy and robustness.

1. Introduction

Cyclones, typhoons, and hurricanes refer to the same meteorological phenomenon in different parts of the world. They are weather systems with strong winds that circulate anticlockwise around a low pressure area in the northern hemisphere and clockwise in the southern hemisphere. Taiwan is located in the northwestern Pacific, on one of the main typhoon paths, and is hit by three to five typhoon events each year on average. However, the rainfall distribution is uneven in both time and space due to the complex terrain conditions in Taiwan. Torrential rain due to typhoons leads to frequent serious disasters such as flooding, landslides, and debris flow. However, the rain is an important water resource that should be stored. Reservoirs are the most important and effective water storage facilities for solving the uneven rainfall

problem. Therefore, reservoir inflow forecasting plays an important role in water resource planning and management.

There are numerous difficulties in constructing a physically based mathematical model because of the extremely complex and highly nonlinear relationship between typhoon rainfall and reservoir inflows. As an attractive alternative to physically based models, data-driven models that are based on artificial intelligence methods, such as neural networks, are favored and are practicably applicable in reservoir inflow forecasting [1–5]. Support vector machines (SVMs) are novel, artificial intelligence-based methods. The SVMs, developed for classification and then extended for regression by Vapnik [6, 7], are based on statistical learning theory. Based on the structural risk minimization (SRM) principle, SVMs theoretically minimize the expected error of a learning machine and reduce the problem of overfitting. In addition,
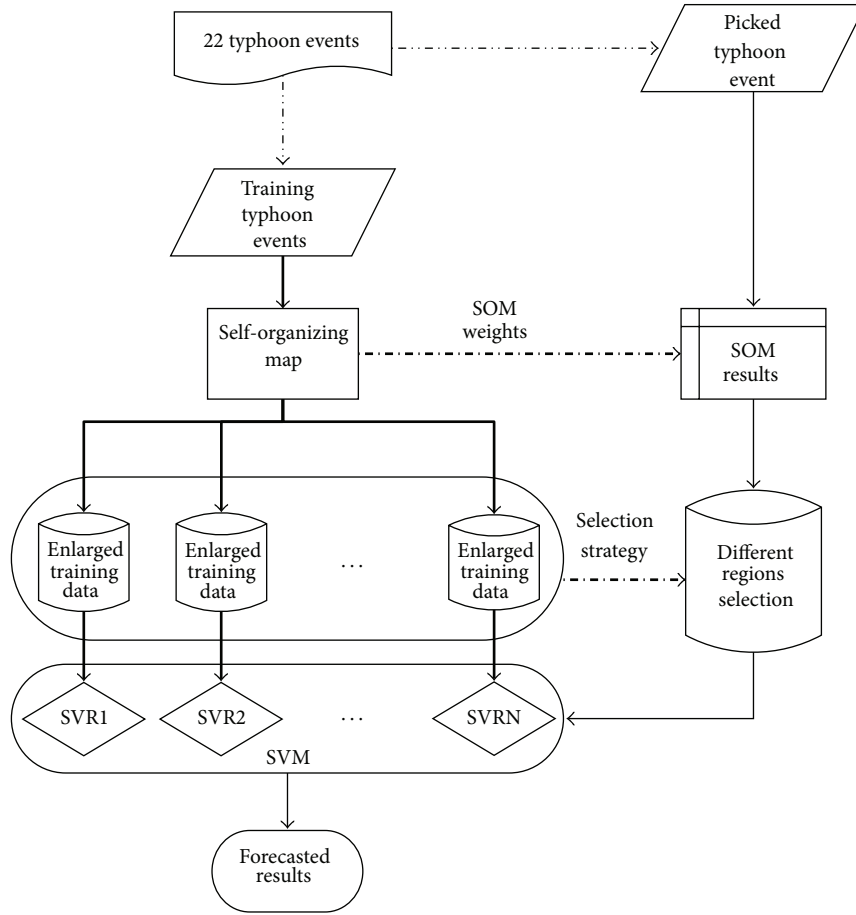
FIGURE 1: Flowchart showing development of the two different SOM-SVM models.

the architecture of the SRM principle guarantees a unique and globally optimal solution by solving the convex optimization problem. A more detailed treatment of SVMs can be found in several text books [8, 9].

The SVMs are proposed as alternative data-driven tools in many fields [10–14] and have excellent generalization ability. In the field of hydrology, problems such as time series forecasting have been reported in recent years [15–28]. The SVMs are outstanding data-driven tools, incorporating the property of regression. However, when the SVM regression considers information with excessive noise or low relationships, the ability of generalization will be reduced. In hydrological cases, it will be found in high values; generalization depends on existence of very many low values. Moreover, for long lead-time forecasting, SVMs can only consider limited noisy information. In such cases, the model cannot forecast the inflow well.

To solve the above problems, in this study, the self-organizing map (SOM) is adopted in advance to group the inputs of SVMs. In each group, the inputs concerned in different inflow processes have a high relationship. Hence, the forecasts obtained by SVMs, which are developed using inputs in the same cluster, may have higher accuracy.

The SOM introduced by Kohonen [29, 30] is a special category of artificial neural networks (ANNs). It can project high-dimensional input space on a low-dimensional topology so as to allow the number of clusters to be determined by inspection. This capability enables the discovery of the relationships among complex data and has been used in recent years [31–40]. Furthermore, the clustering performance of SOM is better than that of conventional clustering methods [41–43].

For improving reservoir inflow forecasting, an approach consisting of a SOM-based clustering method and SVMs is proposed in this study. The SOM density map is obtained using only the past two-hour inflows as inputs for different events and lag time forecasting. Then, SVMs are performed on the basis of the results of the SOM-based clustering method to forecast reservoir inflow. Finally, the proposed approach is applied to the Feitsui Reservoir watershed in northern Taiwan to find the 1 h ahead inflow forecast. A flowchart of the proposed model is illustrated in Figure 1.

2. Methodology

In this paper, the SOM-SVM model, which combines the SOM with SVMs, is proposed. Details about SOM can be found in Kohonen [44]. As to SVMs, one can refer to Vapnik [8] for more details.

2.1. Two SOM-SVM Models. In this study, to avoid the lack of information variability which yields a lower learning ability of the neurons, an enlarged training data set, collected as training forms, is considered. This enlarged training data set implies that besides the training data set of each neuron, all the other training data sets in the same specified region are also adopted as training data sets. To more efficiently obtain well-performed forecasting results, two different specified region strategies are defined. The flowchart for these two specified region strategies concept is shown in Figure 1. The first specified regions are defined by the training data set that, according to the different inflow processes, is SOM-SVM1 model. That is, when the data set on the SOM feature map is simply divided into different training forms, each training form collects the neurons having the same pattern as the different rainfall-runoff processes. The concept of specified region strategy for SOM-SVM1 model is adopted from Hsu et al. [31]. But, different from five regions' definition for Hsu, the specified regions of this study are only four. These training forms, k_j, of the SOM-SVM1 model are denoted by

$$k_j = \{u_x\} \quad j = 1, 2, 3, 4. \tag{1}$$

Here, (1) expresses the xth neurons having the same rainfall-runoff process in the jth region. In general, there are four different rainfall-runoff processes that can be considered specified regions in the typhoon events: increasing inflow region, base flow region, peaking hydrograph region, and recession region. However, for some learning results of the SOM generation, the rainfall-runoff process cannot be clearly divided, especially for the relationships between the increasing inflow region, the peaking hydrograph region, and the recession region. This indicates that, on occasion, there are only three rainfall-runoff training forms found on the SOM feature map. Therefore, when the peaking hydrograph region is not easily separated from the feature map, only three regions can be separated from the SOM feature map.

Different from the SOM-SVM1 specified region definition strategy, the specified regions of the SOM-SVM2 model allow for the selection of stronger relationship neurons. That is, based on the consideration of well-performed results and simple definition of the feature relationship, the enlarged training forms of the SOM-SVM2 model are established. For each neuron generated in the feature map, the SOM-SVM2 model simply collects the enlarged training data sets of some specified higher relation neurons as training forms to implement the SVM model. The concept of the specified higher relation neurons means the neurons in the cross area are adopted as specified higher relation neurons beside themselves. Then, the enlarged training data set is adopted from these specified higher relation neurons. Therefore, for a $n \times n$ SOM feature map, the training forms are denoted by

$$k_j$$
$$= \begin{cases} \{u_j, u_{j+1}, u_{j-n}, u_{j+n}\} & \text{if } u_j = ni + 1 \\ \{u_j, u_{j-1}, u_{j-n}, u_{j+n}\} & \text{if } u_j = ni, \ j = 1, 2, \ldots, n^2 \\ \{u_j, u_{j-1}, u_{j+1}, u_{j-n}, u_{j+n}\} & \text{others,} \end{cases} \tag{2}$$
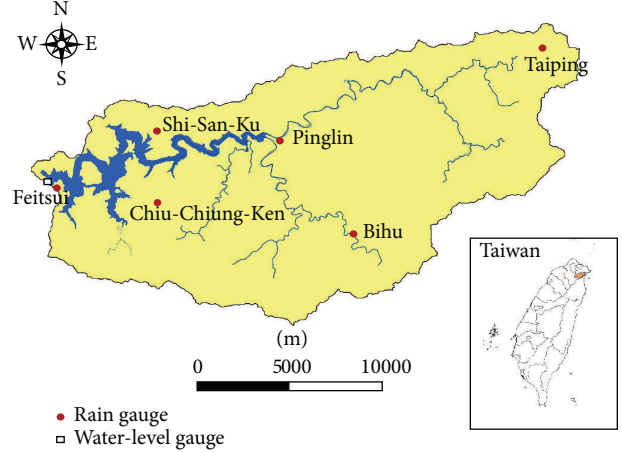

FIGURE 2: The Feitsui Reservoir watershed in northern Taiwan.

where $u_{j-1}, u_{j-n} > 0$, $u_{j+1}, u_{j+n} < n^2$, and the value of i is an integer. When the neurons are located on the feature map, the enlarged training forms have the training forms adopted from the four surrounding neurons and the training forms adopted from itself for each neuron. As the occasions in which $u_{j-1}, u_{j-n} < 0$ or $u_{j+1}, u_{j+n} > n^2$, the neuron should be removed from the equation. It drives for the edge neurons, the enlarged training forms can only adopt the training data sets of the surrounding two or three neurons. This means that these training forms have the training data sets adopted from five neurons, or three or four neurons for the edge neurons. The above training forms generated from two different models are adopted to implement the SVM model. Then, in different specified regions, the SVM models are adopted to generate the forecasts from the forecast forms in each neuron. These forecast forms have only the forecast data sets inside each neuron.

3. Application

3.1. The Study Area and Data. In this paper, all the SVM-based models are applied to the Feitsui Reservoir watershed in northern Taiwan. Feitsui Reservoir is located downstream three major tributaries (Kingkwa Creek, Diyu Creek, and Peishih Creek). It has a surface area of $10 \, \text{km}^2$, a mean depth of 40 m, a maximum depth of 120 m, a full capacity of 406 million m^3, and a total watershed area of $303 \, \text{km}^2$ (see Figure 2). Feitsui Reservoir supplies water for Taipei city (the capital of Taiwan); thus, it is the most important reservoir in northern Taiwan.

The rainfall data are collected from 1988 to 2008. The maximum and average yearly rainfalls are 5736.6 mm and 3808.6 mm, respectively. In this paper, the 22 typhoon events used for model development are presented in Table 1. These 22 typhoon events used for inflow forecasting are divided into two sets, 21 training events and a testing event, in each event's forecasting.

3.1.1. Determination of Lag Length. For the model construction in this paper, it is necessary to decide on the length of

Table 1: The 22 typhoon events used in the modeling.

Name of typhoon	Date (yyyy/mm/dd)	Duration (h)	Peak flow (m^3/s)
Polly	1992/08/27	91	970
Ted	1992/09/20	71	922
Tim	1994/07/09	51	674
Fred	1994/08/19	76	718
Gladys	1994/08/31	34	1,449
Seth	1994/10/08	77	1,457
Herb	1996/07/30	72	2,586
Zeb	1998/10/15	52	2,642
Xangsane	2000/10/31	34	4,075
Nari	2001/09/16	85	4,022
Haiyan	2001/10/15	30	560
Rammasun	2002/07/03	34	1,096
Aere	2004/08/23	70	2,416
Nock-Ten	2004/10/24	46	1,719
Nanmadol	2004/12/03	33	1,690
Haitang	2005/07/17	68	1,476
Matsa	2005/08/04	50	1,226
Talim	2005/08/31	45	1,798
Krosa	2007/10/04	63	2,356
Fung-Wong	2008/07/26	43	969
Sinlaku	2008/09/11	107	3,598
Jangmi	2008/09/26	52	2,513

the forecast form at the beginning. Each forecast form of all the typhoon events is

$$F_{t+\Delta t} = f\left(R_{G,t}, R_{G,t-1}, \ldots, R_{G,t-(L_R-1)}, F_t, F_{t-1}, \ldots,\right.$$
$$\left. F_{t-(L_F-1)}\right), \quad G = 1, 2, \ldots, 7, \tag{3}$$

where t is the current time, Δt is the lead-time period (from 1 to 3 h), $R_{G,t}$ is rainfall in the Gth gauge at time t, F_t is the inflow at time t, f is the derivation of forecasts, and L_F and L_R denote the lag length of inflow and rainfall, respectively.

For the reservoir inflow forecasting model, model construction with appropriate lag lengths of input is an important component. In this paper, the criterion RPE is applied to determine the lag length of inputs. The RPE is defined by

$$\text{RPE} = \frac{E(L) - E(L+1)}{E(L)} \times 100, \tag{4}$$

where $E(L)$ and $E(L+1)$ are the root-mean-square-error (RMSE) for the model with L and $L+1$ lag lengths, respectively. The RMSE can be obtained as

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left(\widehat{F}_t - F_t\right)^2}, \tag{5}$$

where F_t is the inflow at time t, $\widehat{F}_t$ is the predicted inflow at time t, and n is the number of time steps. In general, the RMSE

decreases with increasing lag term. When the RPE value is less than 5%, the increase of lag lengths is stopped.

By using this procedure, the most appropriate lag lengths of typhoon rainfall, L_R, and inflow, L_F, for a certain lead-time, Δt, can be determined. The appropriate lag lengths for two hours of rainfall and inflow are used to forecast the 1- to 3-hour ahead inflows. Then, the general form with appropriate lag lengths for the SVM-based models is described as

$$F_{t+\Delta t} = f\left(R_{G,t-1}, R_{G,t-2}, F_{t-1}, F_{t-2}\right) \quad G = 1, 2, \ldots, 7. \tag{6}$$

Additionally, for reasonable model comparison, the two indices of the SVM regression, C and ε, are simply defined as 1 and 0.1.

3.1.2. Criteria. To discuss the individual and average performance of the SVM and two different SOM-SVM based models, three indices, relative root-mean-square-error (RRMSE), mean root-mean-square-error (MRMSE), and mean coefficient of efficiency (MCE), are used:

(1) Relative root-mean-square-error (RRMSE):

$$\text{RRMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left(\frac{\widehat{F}_t - F_t}{F_t}\right)^2}. \tag{7}$$

(2) Mean root-mean-square-error (MRMSE):

$$\text{MRMSE} = \frac{1}{N} \sum_{i=1}^{l} \text{RMSE}_i. \tag{8}$$

(3) Mean coefficient of efficiency (MCE):

$$\text{CE} = 1 - \frac{\sum_{t=1}^{n} \left(F_t - \widehat{F}_t\right)^2}{\sum_{t=1}^{n} \left(F_t - \overline{F}_t\right)^2},$$
$$\text{MCE} = \frac{1}{N} \sum_{i=1}^{l} \text{CE}_i, \tag{9}$$

where $\overline{F}_t$ is the average of observed inflows and l is the number of forecasting typhoon events.

3.2. SOM-SVM Models. As a comparison to strengthen the SVM-based model ability, the forecasts of the original SVM model are directly generated from all the forecasting data sets (where the SVM regressions are built from the other 21 typhoon events). Then, in contrast to the original SVM model, the SOM clustering and SVMs regression are combined for both of the proposed SVM-based models. Initially, all forecast forms in each event are divided into different neurons in the SOM clustering process from which the SOMs are established by the forecast forms of the other typhoon events. Then, the forecast forms divided for these different SOM neurons are used to generate forecasts by different strategies. After the forecasts are obtained by SVMs regression, the forecasts are combined according to the original time series

FIGURE 3: The ID number and location of each neuron for a SOM with 5×5 dimensions.

FIGURE 4: The density map resulting from the SOM of 5×5 dimensions according to the 21 typhoon events (taking Typhoon Polly as the testing event).

process. The difference between the SOM-SVM1 and SOM-SVM2 models is the training form selection strategy, which is mentioned above. To highlight the advantage of the selection strategy of the SOM-SVM1 model, the original SVM forecasts are generated for comparison with the performance of the SOM-SVM1 model. Furthermore, based on the concept of clusters considering the lead-time information, it may make the forecasts not able to obtain some important information from the neighbor neurons for different regions in its cluster strategy. The selection strategy of the SOM-SVM2 model that considers higher relationship inputs is then established. The performance of these three SVM-based models is finally estimated to allow for intercomparison.

3.2.1. SOM Clustering. In this subsection, a SOM with a small dimension is considered the best option. If the clustering result is reasonable and satisfactory, the cluster analysis can be accepted. Otherwise, another SOM with a larger dimension is chosen to analyze input patterns. This step is repeated until a satisfactory result is obtained. That is, the inputs within each grid have the same characteristics associated with a certain inflow process. Then, for different events in different lead-time forecasts, SOMs are generated from the same process. Taking the 1 h ahead forecasts of the Polly Typhoon event for instance, the SOM is constructed from the other 21 events. According to our experiments, a 5×5 dimension SOM is adopted herein. That is, the competitive layer contains 25 grids. After the SOM clustering is implemented, the corresponding feature map and density map can be obtained.

In the feature map, forecast forms with similar rainfall-runoff processes are located in neighboring grids. On the other hand, forecast forms with significantly different rainfall-runoff processes are located in different grids that are distant from each other. Such a characteristic is also retained in the density map, because the density map results from the feature map. The ID number and location of the 5×5 SOM generated from the 21 typhoon events (except for the Polly Typhoon) are presented in Figure 3. In addition, its density map is shown in Figure 4. In Figure 4, the number inside each neuron indicates how many training data sets are projected onto the same topology point.

FIGURE 5: The classification results for the SOM-SVM1 model according to the 21 typhoon events (taking Typhoon Polly as the testing event).

3.2.2. SOM-SVM1 Model. According to the above SOM results, it is found that the map can be divided into four specified regions with the rainfall-runoff process characteristics of different neurons. The specified region divisions are shown in Figure 5. The association between the four specified regions and the rainfall-runoff processes is also clearly illustrated in Figure 6. The regions are (1) increasing inflow region (region 1), (2) base flow region (region 2), (3) peaking hydrograph region (region 3), and (4) recession region (region 4). For example, region 1, located in the upper left area of the SOM (Figure 5), represents increasing reservoir inflow during the period (Figure 6). In this clustering process, all the forecast forms of each event are divided into different neurons. Then, according to (1), the forecasting pattern within the neurons can be expressed as

$$\left(F_{t+\Delta t}\right)_{k_j} = f\left(R_{G,t}, R_{G,t-1}, F_t, F_{t-1}\right)_{u_x}, \qquad (10)$$

where, as in the equation mentioned above, k_j expresses the jth specified region and u_x expresses the xth neuron.
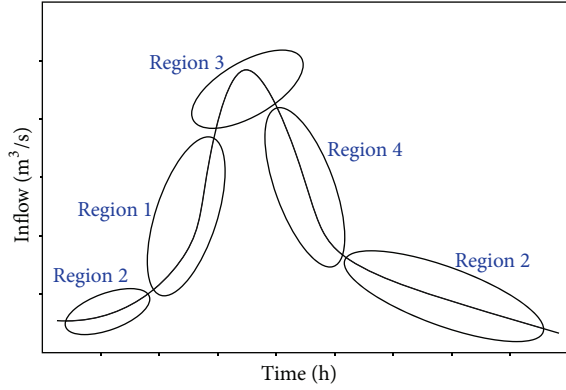
FIGURE 6: The relationship between the four classification regions and the rainfall-runoff processes for the SOM-SVM1 model.

After the above process, 22 SOMs could be generated, and each SOM is generated from the other 21 typhoon events. Among all 22 events, the longest event is 107 h, and the shortest event is 33 h; therefore, the SOMs are not generated from the same date length and data sets. Consequently, each generated SOM may be divided into four specified regions but is always a little different from the other 21 maps. The training forms of each neuron in the same specified region are collected as training form sets for model establishment in different specified regions. These results are used for SVMs regression to obtain the forecast forms of each neuron in the same specified region. Thus, the model SOM-SVM1 is established.

3.2.3. SOM-SVM2 Model. The above SOM-SVM1 model is established considering high relativity from four reservoir inflow processes. However, some training forms might be allocated to different neurons and, thus, further allocated to different specified regions for different events. This means that these training forms sometimes still have important information for the other specified regions but cannot pass this information to them. Besides, with the different strengths and lengths of the 22 typhoon events, the training forms of each typhoon event would not distribute to the 25 neurons on the feature map evenly. The maximum density number for the 5×5 feature map is 295, and the minimum number is 12. For different flow processes, the high density map may have less information for other neurons, while the low density map may have more information. However, a large amount of information can be found for close neurons. Furthermore, as the clusters consider lead-time information, some of the forecasting forms of the SOM-SVM1 model cannot pass information to the forecasts in different regions. To cope with these outcomes, the SOM-SVM2 model is proposed.

According to the feature map generated from the SOM discussed in Section 3.2.1, a different enlarged training form is simply adopted here. The concept of this enlarged training form implies that, except for the training data set of each neuron, only the training data sets of surrounding neurons in the cross region are adopted as training forms. This means that, in the cross region of each neuron, four is the maximum number of surrounding neurons containing forecast forms that can be adopted as training forms for each neuron. However, for the neurons located on the edge of the competitive layer, only two or three surrounding neurons are adopted as training forms. Then, according to (2), the forecasting form of a $n \times n$ SOM feature map can be written as

$$(F_{t+\Delta t})_{k_j} = \begin{cases} f\left(R_{G,t-1}, R_{G,t-2}, F_{t-1}, F_{t-2}\right)_{u_j, u_{j+1}, u_{j-n}, u_{j+n}} & \text{if } u_j = ni + 1 \\ f\left(R_{G,t-1}, R_{G,t-2}, F_{t-1}, F_{t-2}\right)_{u_j, u_{j-1}, u_{j-n}, u_{j+n}} & \text{if } u_j = ni, \; j = 1, 2, \ldots, 25 \\ f\left(R_{G,t-1}, R_{G,t-2}, F_{t-1}, F_{t-2}\right)_{u_j, u_{j-1}, u_{j+1}, u_{j-n}, u_{j+n}} & \text{others,} \end{cases} \tag{11}$$

where $u_{j-1}, u_{j-n} > 0$, $u_{j+1}, u_{j+n} < n^2$, j expresses the neuron ID number, and i expresses integers. Table 2 presents the lists of neuron ID numbers for the enlarged training data set chosen from different neurons for the Polly Typhoon event generated by the SOM-SVM2 model. Taking the neuron ID numbers 1, 10, and 17 listed in Table 2 as examples, the enlarged training data set of these neurons is presented in Figure 7. The same enlarged training data set generations are produced within the same structure for each typhoon event. Then, the forecast forms are adopted to implement the SVMs for the SOM-SVM2 model results generation. Different from the SOM-SVM1 model, the enlarged training data set of SOM-SVM2 is not selected by considering the inflow process. In this study, irrespective of whether the number of neurons is selected, the training forms are simply selected from the neuron location. This enlarged training form selection strategy not only considers the training data with higher relationships but also does not require to exclude the training data set in a different specified region. Similarly, the above forecast forms are adopted to implement the SVMs for inflow forecasting.

3.2.4. Comparison between the Original SVM and the SOM-SVM1. In the SOM-SVM1 model, the SOM process clusters the forecast forms into different neurons according to the input characteristics and then divides them into different regions related to different inflow processes of the typhoon events. It follows that the forecasts can be generated considering high relative inputs, and the lower relative inputs can be ignored. Taking Typhoon Polly, which is the typhoon event with different rise and fall rainfall-runoff processes, as an example, the SOM-SVM1 model can

TABLE 2: The neurons containing the testing data set (1st column) and the neurons containing the enlarged training data set (2nd column), for the SOM-SVM2 model.

Neuron ID number of typhoon a	Neuron ID number of typhoons $i = 1, \ldots, 22$, $i \neq a$
1	1, 2, 6
2	1, 2, 3, 7
3	2, 3, 4, 8
4	3, 4, 5, 9
5	4, 5, 10
6	1, 6, 7, 11
7	2, 6, 7, 8, 12
8	3, 7, 8, 9, 13
9	3, 7, 8, 9, 13
10	5, 9, 10, 15
11	6, 11, 12, 16
12	7, 11, 12, 13, 17
13	8, 12, 13, 14, 18
14	9, 13, 14, 15, 19
15	10, 14, 15, 20
16	11, 16, 17, 21
17	12, 16, 17, 18, 22
18	13, 17, 18, 19, 23
19	14, 18, 19, 20, 24
20	15, 19, 20, 25
21	16, 21, 22
22	17, 21, 22, 23
23	18, 22, 23, 24
24	19, 23, 24, 25
25	20, 24, 25

Note: please refer to Figure 7 for the ID numbers and locations of neurons.
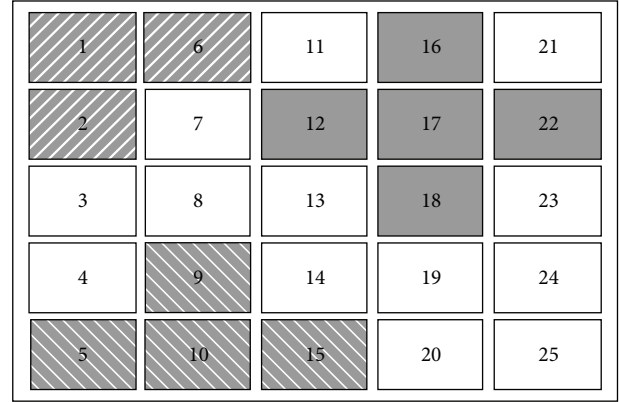

FIGURE 7: The neurons containing the enlarged training data set for the SOM-SVM2 model (taking neuron ID numbers 1, 10, and 17 as examples).

TABLE 3: Performance comparison of the SVM, SOM-SVM1, and SOM-SVM2 models for 1 h ahead forecasts using RRMSE and CE as criteria.

Typhoon events	SOM		SOM-SVM1		SOM-SVM2	
	RRMSE	CE	RRMSE	CE	RRMSE	CE
Polly	1.358	0.583	0.448	0.912	0.374	0.916
Ted	0.316	0.770	0.157	0.941	0.128	0.958
Tim	3.627	0.551	1.724	0.873	1.253	0.901
Fred	1.460	0.745	0.369	0.960	0.365	0.952
Gladys	7.430	0.644	1.354	0.921	0.545	0.917
Seth	5.882	0.650	0.817	0.936	0.731	0.950
Herb	1.375	0.959	0.367	0.967	0.268	0.975
Zeb	0.340	0.905	0.187	0.958	0.186	0.955
Xangsane	0.280	0.918	0.187	0.915	0.151	0.936
Nari	0.268	0.808	0.263	0.895	0.222	0.925
Haiyan	1.222	0.743	0.471	0.912	0.337	0.940
Rammasun	2.975	0.643	0.740	0.929	0.613	0.929
Aere	2.784	0.935	1.245	0.979	1.057	0.988
Nock-Ten	0.919	0.788	0.769	0.961	0.657	0.937
Nanmadol	1.681	0.806	0.258	0.924	0.237	0.908
Haitang	7.049	0.727	2.329	0.923	1.225	0.948
Matsa	4.292	0.523	1.474	0.904	1.032	0.900
Talim	3.623	0.874	0.874	0.883	0.779	0.958
Krosa	0.667	0.960	0.554	0.980	0.402	0.981
Fung-Wong	0.876	0.830	0.286	0.979	0.262	0.980
Sinlaku	0.696	0.858	0.212	0.939	0.226	0.913
Jangmi	0.400	0.937	0.133	0.986	0.132	0.978

generate well-performed forecasts by considering forecast forms belonging to one of the inflow specified regions. Compared to the forecasts generated from the original SVM model, the original SVM generates the forecasts irrespective of whether or not the forecast forms have extreme values. However, the forecasts generated from the SOM-SVM1 model can arrange the forecast forms belonging to different rainfall-runoff processes. That is, the SOM-SVM1 model is established to strengthen the forecast ability for different rainfall-runoff processes.

The inflow results generated from the original SVM model and SOM-SVM1 model for 1 h ahead forecasting are arranged in Table 3. These show that, compared to the forecasts generated from the original SVM model, the RRMSE values derived from the SOM-SVM1 model are small for each typhoon event. Even the original SVM model can derive a better CE value for the Xangsane typhoon. This simply means that the CE value of the original SVM can perform better only for the Xangsane typhoon, and both models can generate a CE value up to 0.9. Actually, the CE values generated from

the original SVM model only have six events reaching 0.9, while the SOM-SVM1 model has 19. In addition, according to Figures 8(a) and 8(b), it can be found that the SOM-SVM1 model generates forecasts with significantly less scatter than the original SVM model when plotted against the measured values. Further lead-time results are presented in Table 4. It can be found that the average values of both the CE value
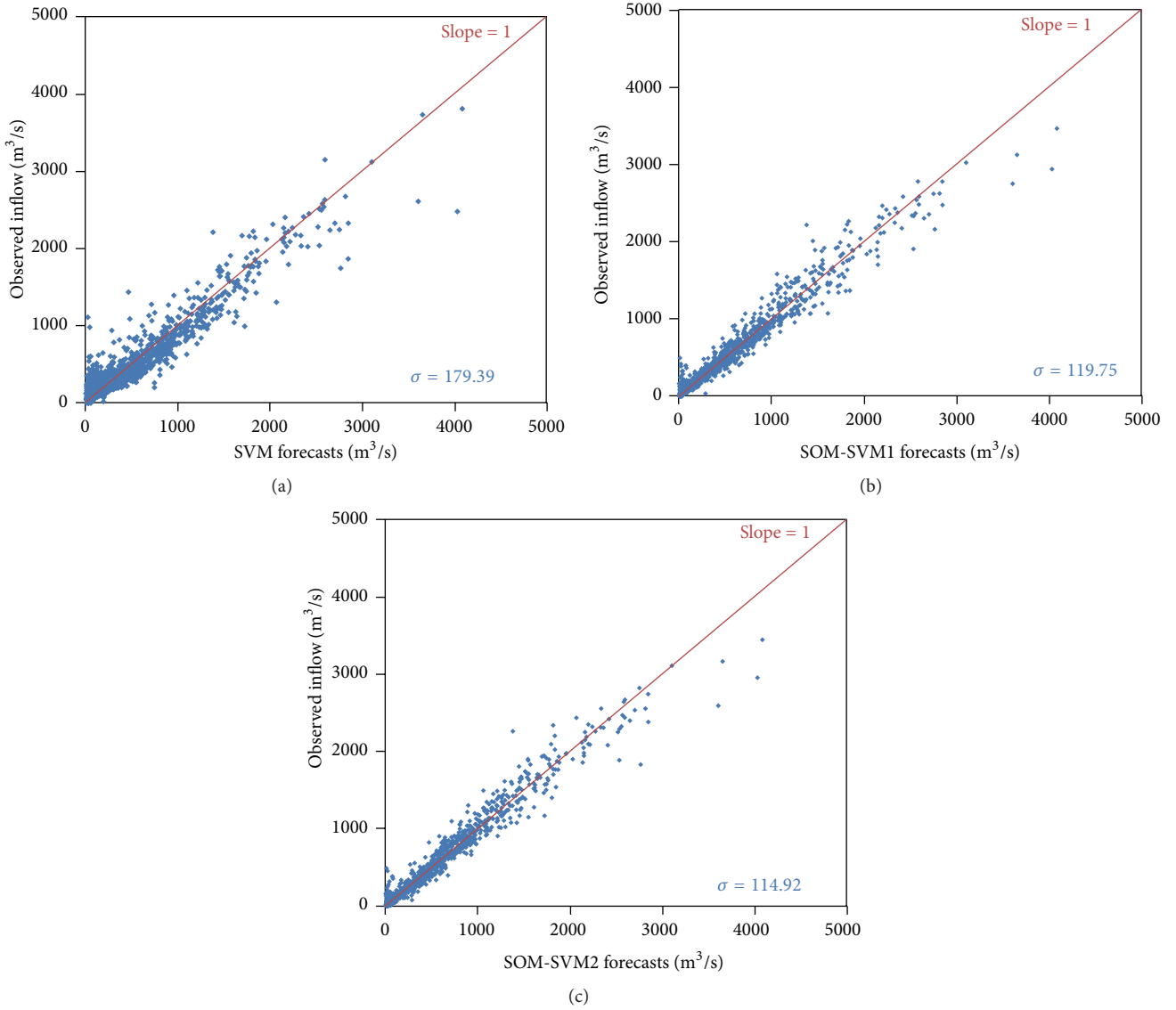
(a)

(b)

(c)

FIGURE 8: The observed inflows versus the forecasts obtained by (a) the original SVM model, (b) the SOM-SVM1 model, and (c) the SOM-SVM2 model.

TABLE 4: Performance comparison of SVM and SOM-SVM for different lead-times using MRMSE and MCE as criteria.

Lead-time (h)	MRMSE (m^3/s)			MCE		
	SVM	SOM-SVM1	SOM-SVM2	SVM	SOM-SVM1	SOM-SVM2
1	167.078	99.535	96.899	0.780	0.934	0.943
2	198.373	156.716	144.010	0.729	0.864	0.884
3	252.267	229.102	204.552	0.560	0.690	0.759

and RRMSE values of the SOM-SVM1 model for different lead-times are still better than the original SVM model. The improvement of both indices are 21.5%, 18.4%, and 23.0% (arranged in Table 5), respectively.

3.2.5. SOM-SVM Results Comparison. All three SVM-based models are compared in this section. Taking Typhoon Polly as an example, a comparison of the three different SVM-based

reservoir inflow forecasts for 1 h ahead to 3 h ahead is presented in Figures 9–11. In these figures, the forecasts generated from the SOM-SVM2 model performed better than the original SVM model but did not perform markedly better than the SOM-SVM1 model. This is because, after the SOM cluster process, all the forecast forms with similar input characteristics are arranged to closer, or the same, neurons. As the SVM-based models are established from these

TABLE 5: The improvement in MCE due to the use of SOM-SVM2 and SOM-SVM1 instead of SVM.

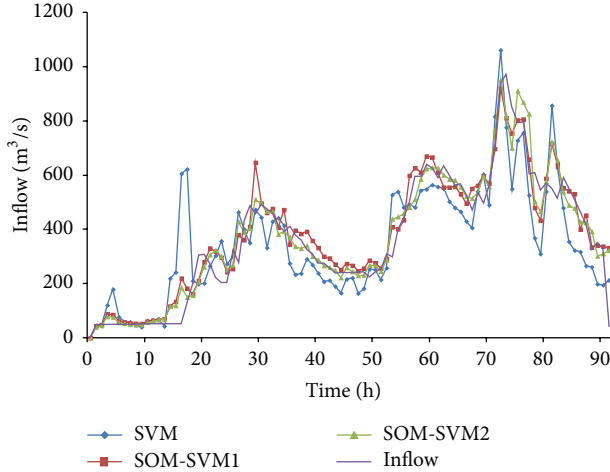| Lead-time (h)	Improvement (%)	
	SOM-SVM1	SOM-SVM2
1	19.7	20.9
2	18.5	21.2
3	23.0	35.4

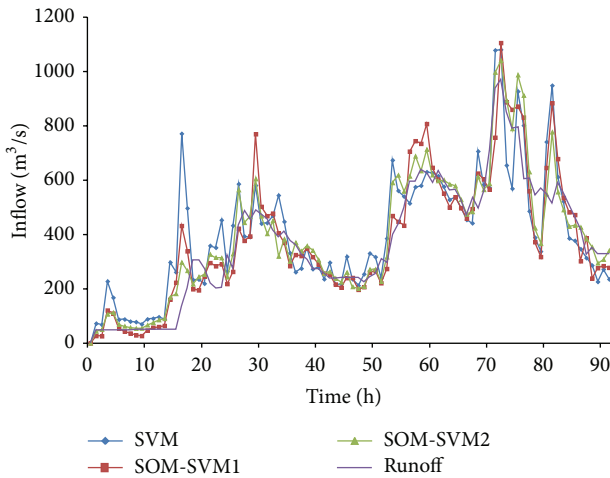FIGURE 9: Comparison of the observed inflow with the 1 h ahead forecasts for Typhoon Polly (1 h).

FIGURE 10: Comparison of the observed inflow with the 1 h ahead forecasts for Typhoon Polly (2 h).

FIGURE 11: Comparison of the observed inflow with the 1 h ahead forecasts for Typhoon Polly (3 h).

FIGURE 12: The MCE values of the three SVM-based models.

arranged inputs (selected from closer meaningful neurons) instead of all the inputs, the models can be strengthened for the relative selection strategies. The purpose of this study is to find an excellent selection strategy for typhoon events. Figures 8(a), 8(b), and 8(c) are again compared here. In Figures 8(a) and 8(b), the SOM-SVM1 models can generate higher concentrated forecasts, as mentioned above. However, in Figure 8(c), it can still be found that the forecasts of the SOM-SVM2 model have higher concentration than the other two SVM-based models.
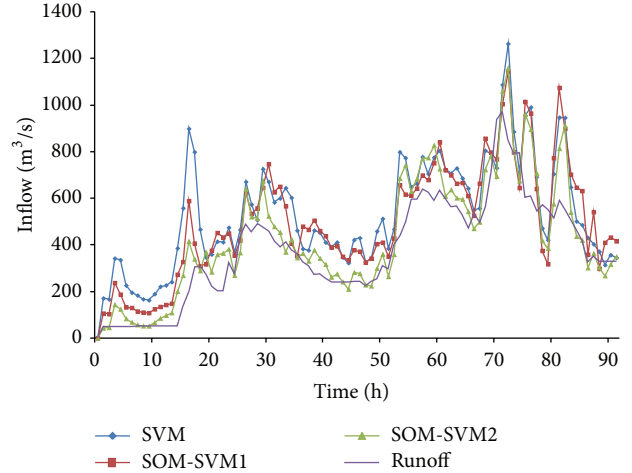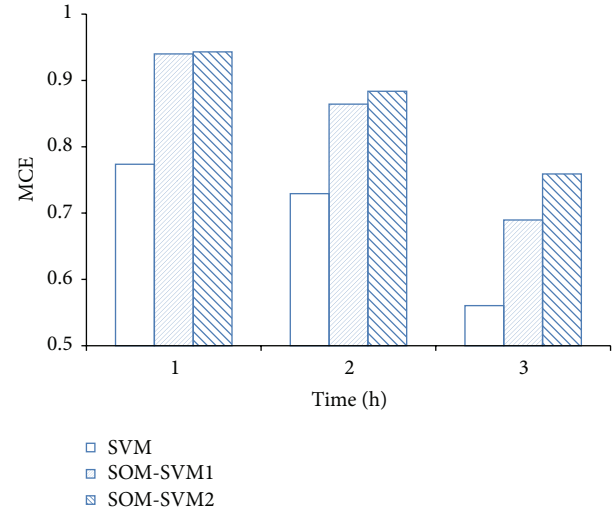
All three SVM-based model forecasting results for 1 h ahead are shown in Table 4. It can be found that the SOM-SVM1 model is outperforming the original SVM model, as mentioned above. However, except for the fact that all RRMSE values for the SOM-SVM2 model are lower than both of the original SVM model and the SOM-SVM1 model, the SOM-SVM2 model does not perform better than the SOM-SVM1 model at each CE value of the typhoon events. Nevertheless, in contrast to the other two SVM-based models, all the CE values of SOM-SVM2 surpass 0.9. Table 5 lists the improvements to MCE computed from the SOM-SVM1 and SOM-SVM2 models when compared with the original SVM model. The same improvements to MCE values are shown in Figure 12. These show that the SOM-SVM2 model can generate a more than 20% improvement over the original SVM model, and this improvement reaches 35.4% in 3 h ahead forecasts. This is excellent when compared to the other two models.

Table 6: The improvement in MCE due to the use of SOM-SVM2 instead of SOM-SVM1.

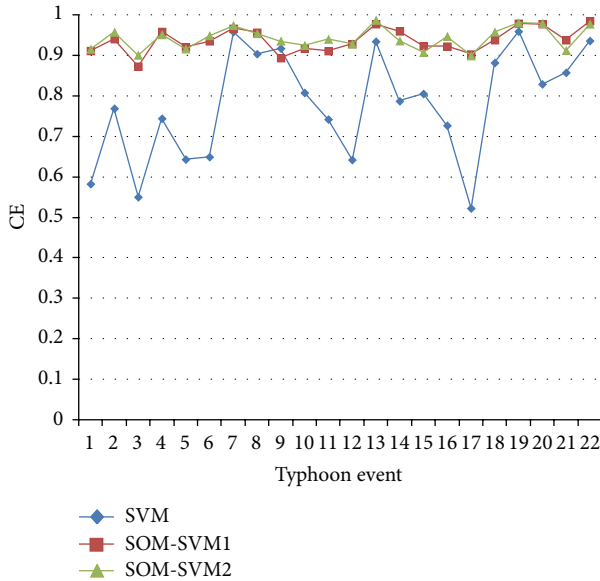| Lead-time (h)	Improvement (%)
1	0.96
2	2.25
3	10.08

Figure 13: The CE values for 1 h ahead forecasts from three SVM-based models.

Finally, the performances of the SOM-SVM1 and SOM-SVM2 models are compared again. As mentioned above, all the CE values of the SOM-SVM2 model exceed 0.9, but only 19 values exceed 0.9 for the SOM-SVM1 in 1 h forecasts. However, this does not mean that the quantity of CE values in 1 h ahead forecasts for the SOM-SVM2 model is higher than for the SOM-SVM1 model in every typhoon event. As the list in Table 4 shows, the SOM-SVM1 model has the better CE values for 8 events within all 22 events. However, the MCE indicates that the average CE value of the SOM-SVM2 model is better than for the SOM-SVM1 model. In any case, the RRMSE values still show that the SOM-SVM2 model can generate less inaccurate forecasts, as shown in Table 4.

Comparing the SOM-SVM1 and SOM-SVM2 models, the improvements for the SOM-SVM2 model are arranged in Table 6 and Figure 13. The MCE value generated from the SOM-SVM2 model is only 0.009 greater than from the SOM-SVM1 model for 1 h ahead forecasts. This means the improvement of the SOM-SVM2 model over the SOM-SVM1 model is just 0.96%. However, for 2 h and 3 h ahead inflow forecasts, the SOM-SVM2 model obviously improved, with 2.25% and 10.08% improvements, respectively. This means that after the training data set selected from the SOM-SVM2 model is adopted, the forecasts can be better generated than for the training data set selected from the SOM-SVM1 model. As the time is extended, the performance of SOM-SVM2

shows an strong improvement compared to the two other models.

Briefly, the model SOM-SVM2 is built to forecast typhoon events with simple structure concept and limited higher relative data. For the SOM-SVM1 model, the data distribution on the SOMs should be clustered for considering different reservoir inflow processes in advance and it makes the forecasts of typhoon events be generated from fewer and higher relative data. However, the SOM-SVM2 model is robust for the cross region selection strategy does not need to understand the distribution on the SOMs for reservoir inflow processes and still can collect fewer and higher relative data. Furthermore, the SOM-SVM2 model can generate better performance of typhoon events forecasts than both the original SVM model and SOM-SVM1 model.

4. Summary and Conclusions

The objective of this paper is to develop a precise and stable reservoir inflow forecasting model for reservoir operations during typhoon periods. For this purpose, instead of the original SVM model, two different enlarged training form selection strategies from SOM are combined to construct a piecewise nonlinear model. The first is the model that considers inputs selected from the inflow processes: the SOM-SVM1 model. The second is the model adopted for the training form of neurons on the SOM feature map: the SOM-SVM2 model.

In conclusion, there are at least three reasons favoring the use of the SOM-SVM2 model for inflow forecasts. First, both of the developed SOM-combined SVM models are established from the SOM model that strengthens the forecast ability well. Second, the SOM-SVM2 model can adopt the forecasting forms without considering the clusters on the SOM feature map. For the models established, the SOM-SVM2 model needs an enlarged training form containing the training data set of each of the neurons and the surrounding neurons. These neurons are selected depending only on the situation of the SOM feature map without external inflow definition for higher relationships. Finally, although the SOM-SVM2 model only derives a 0.96% improvement in 1 h reservoir inflow forecast results, it derives 2.25% and 10.08% improvements in MCE value for 2 h and 3 h lead-time forecast results, respectively, when compared to the SOM-SVM1 model. Moreover, the SOM-SVM2 model exhibits 20.9%, 21.2%, and 34.4% improvements for the MCE values for reservoir inflow forecast results with 1 h, 2 h, and 3 h lead-times, respectively, when compared to the original SVM model. In other words, the advantage of the SOM-SVM2 model becomes most obvious in long-term forecasts. The proposed SOM-SVM2 model is recommended as an alternative to the existing models because of its accuracy, robustness, and efficiency. This modeling technique is expected to be useful in improving reservoir inflow forecasting.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] P. Coulibaly, F. Anctil, and B. Bobée, "Daily reservoir inflow forecasting using artificial neural networks with stopped training approach," *Journal of Hydrology*, vol. 230, no. 3-4, pp. 244–257, 2000.

[2] G.-F. Lin and L.-H. Chen, "A non-linear rainfall-runoff model using radial basis function network," *Journal of Hydrology*, vol. 289, no. 1–4, pp. 1–8, 2004.

[3] Y.-H. Chen and F.-J. Chang, "Evolutionary artificial neural networks for hydrological systems forecasting," *Journal of Hydrology*, vol. 367, no. 1-2, pp. 125–137, 2009.

[4] G.-F. Lin, M.-C. Wu, G.-R. Chen, and F.-Y. Tsai, "An RBF-based model with an information processor for forecasting hourly reservoir inflow during typhoons," *Hydrological Processes*, vol. 23, no. 25, pp. 3598–3609, 2009.

[5] G.-F. Lin and M.-C. Wu, "An RBF network with a two-step learning algorithm for developing a reservoir inflow forecasting model," *Journal of Hydrology*, vol. 405, no. 3-4, pp. 439–450, 2011.

[6] V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, NY, USA, 1998.

[7] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999.

[8] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.

[9] N. Cristianini and J. Shaw-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, New York, NY, USA, 2000.

[10] F. Bovolo, L. Bruzzone, and M. Marconcini, "A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 7, pp. 2070–2082, 2008.

[11] B. Y. Hiew, A. B. J. Teoh, and O. S. Yin, "A secure digital camera based fingerprint verification system," *Journal of Visual Communication and Image Representation*, vol. 21, no. 3, pp. 219–231, 2010.

[12] X. Song and A. M. Wyrwicz, "Unsupervised spatiotemporal fMRI data analysis using support vector machines," *NeuroImage*, vol. 47, no. 1, pp. 204–212, 2009.

[13] A. Baylar and M. Batan, "Usage of artificial intelligence methods in free flowing gated closed conduits for estimation of oxygen transfer efficiency," *Advances in Engineering Software*, vol. 41, no. 5, pp. 729–736, 2010.

[14] A. Muñoz and J. González, "Representing functional data using support vector machines," *Pattern Recognition Letters*, vol. 31, no. 6, pp. 511–516, 2010.

[15] S.-Y. Liong and C. Sivapragasam, "Flood stage forecasting with support vector machines," *Journal of the American Water Resources Association*, vol. 38, no. 1, pp. 173–186, 2002.

[16] X. Y. Yu, S. Y. Liong, and V. Babovic, "EC-SVM approach for real-time hydrologic forecasting," *Journal of Hydroinformatics*, vol. 6, no. 3, pp. 209–223, 2004.

[17] C. Sivapragasam and S.-Y. Liong, "Flow categorization model for improving forecasting," *Nordic Hydrology*, vol. 36, no. 1, pp. 37–48, 2005.

[18] P.-S. Yu, S.-T. Chen, and I.-F. Chang, "Support vector regression for real-time flood stage forecasting," *Journal of Hydrology*, vol. 328, no. 3-4, pp. 704–716, 2006.

[19] X. Y. Yu and S.-Y. Liong, "Forecasting of hydrologic time series with ridge regression in feature space," *Journal of Hydrology*, vol. 332, no. 3-4, pp. 290–302, 2007.

[20] T.-Y. Pan and R.-Y. Wang, "State space neural networks for short term rainfall-runoff forecasting," *Journal of Hydrology*, vol. 297, no. 1-4, pp. 34–50, 2004.

[21] S.-T. Chen and P.-S. Yu, "Pruning of support vector networks on flood forecasting," *Journal of Hydrology*, vol. 347, no. 1-2, pp. 67–78, 2007.

[22] G.-F. Lin, G.-R. Chen, M.-C. Wu, and Y.-C. Chou, "Effective forecasting of hourly typhoon rainfall using support vector machines," *Water Resources Research*, vol. 45, no. 8, 2009.

[23] G.-F. Lin, G.-R. Chen, and P.-Y. Huang, "Effective typhoon characteristics and their effects on hourly reservoir inflow forecasting," *Advances in Water Resources*, vol. 33, no. 8, pp. 887–898, 2010.

[24] H. P. Tserng, G.-F. Lin, L. K. Tsai, and P.-C. Chen, "An enforced support vector machine model for construction contractor default prediction," *Automation in Construction*, vol. 20, no. 8, pp. 1242–1249, 2011.

[25] G.-F. Lin, Y.-C. Chou, and M.-C. Wu, "Typhoon flood forecasting using integrated two-stage support vector machine approach," *Journal of Hydrology*, vol. 486, pp. 334–342, 2013.

[26] G.-F. Lin, P.-K. Huang, and H.-Y. Lin, "Forecasting tropical cyclone intensity change in the western North Pacific," *Journal of Hydroinformatics*, vol. 15, no. 3, pp. 952–966, 2013.

[27] G.-F. Lin, B.-C. Jhong, and C.-C. Chang, "Development of an effective data-driven model for hourly typhoon rainfall forecasting," *Journal of Hydrology*, vol. 495, pp. 52–63, 2013.

[28] M.-C. Wu, G.-F. Lin, and H.-Y. Lin, "Improving the forecasts of extreme streamflow by support vector regression with the data extracted by self-organizing map," *Hydrological Processes*, vol. 28, no. 2, pp. 386–397, 2014.

[29] T. Kohonen, "The self organizing map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.

[30] T. Kohonen, *Self-Organizing Maps*, vol. 30 of *Springer Series in Information Sciences*, Springer, Berlin, Germany, 1995.

[31] K.-L. Hsu, H. V. Gupta, X. Gao, S. Sorooshian, and B. Imam, "Self-organizing linear output map (SOLO): an artificial neural network suitable for hydrologic modeling and analysis," *Water Resources Research*, vol. 38, no. 12, pp. 38-1–38-17, 2002.

[32] H. Moradkhani, K.-L. Hsu, H. V. Gupta, and S. Sorooshian, "Improved streamflow forecasting using self-organizing radial basis function artificial neural networks," *Journal of Hydrology*, vol. 295, no. 1–4, pp. 246–262, 2004.

[33] G.-F. Lin and L.-H. Chen, "Time series forecasting by combining the radial basis function network and the self-organizing map," *Hydrological Processes*, vol. 19, no. 10, pp. 1925–1937, 2005.

[34] H. Yang, K. L. Hsu, S. Sorooshian, and X. Gao, "Self-organizing nonlinear output (SONO): a neural network suitable for cloud patch-based rainfall estimation from satellite imagery at small scales," *Water Resources Research*, vol. 41, Article ID W03008, 2005.

[35] E. V. Samsonova, J. N. Kok, and A. P. Ijzerman, "TreeSOM: cluster analysis in the self-organizing map," *Neural Networks*, vol. 19, no. 6-7, pp. 935–949, 2006.

[36] S. Srinivasulu and A. Jain, "A comparative analysis of training methods for artificial neural network rainfall-runoff models," *Applied Soft Computing*, vol. 6, no. 3, pp. 295–306, 2006.

[37] G.-F. Lin and C.-M. Wang, "Performing cluster analysis and discrimination analysis of hydrological factors in one step," *Advances in Water Resources*, vol. 29, no. 11, pp. 1573–1585, 2006.

[38] G.-F. Lin and M.-C. Wu, "A SOM-based approach to estimating design hyetographs of ungauged sites," *Journal of Hydrology*, vol. 339, no. 3-4, pp. 216–226, 2007.

[39] M. Alvarez-Guerra, C. González-Piñuela, A. Andrés, B. Galán, and J. R. Viguri, "Assessment of Self-Organizing Map artificial neural networks for the classification of sediment quality," *Environment International*, vol. 34, no. 6, pp. 782–790, 2008.

[40] D. Müller, "Self organized mapping of data clusters to neuron groups," *Neural Networks*, vol. 22, no. 4, pp. 415–424, 2009.

[41] P. Mangiameli, S. K. Chen, and D. West, "A comparison of SOM neural network and hierarchical clustering methods," *European Journal of Operational Research*, vol. 93, no. 2, pp. 402–417, 1996.

[42] S. C. Michaelides, C. S. Pattichis, and G. Kleovoulou, "Classification of rainfall variability by using artificial neural networks," *International Journal of Climatology*, vol. 21, no. 11, pp. 1401–1414, 2001.

[43] G.-F. Lin and L.-H. Chen, "Identification of homogeneous regions for regional frequency analysis using the self-organizing map," *Journal of Hydrology*, vol. 324, no. 1–4, pp. 1–9, 2006.

[44] T. Kohonen, *Self-Organizing Maps*, Springer, New York, NY, USA, 2001.

Journal of
Geochemistry

Scientifica

International Journal of
Ecology

The Scientific
World Journal

Journal of
Mining

Journal of
Earthquakes

Paleontology Journal

Journal of
Petroleum Engineering

Hindawi

Submit your manuscripts at
http://www.hindawi.com

International Journal of
Geophysics

Advances in
Meteorology

Journal of
Climatology

Advances in
Geology

Advances in
Oceanography

International Journal of
Oceanography

International Journal of
Mineralogy

Journal of
Geological Research

International Journal of
Atmospheric Sciences

Applied &
Environmental
Soil Science

Journal of
Computational
Environmental Sciences